

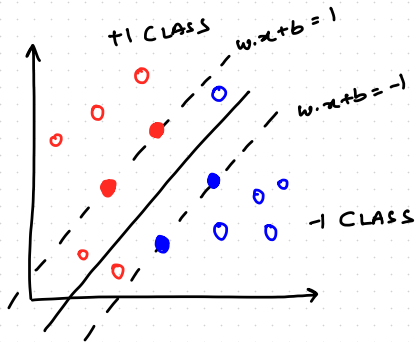
SVM Soft Margin Classification

Nipun Batra

IIT Gandhinagar

August 1, 2025

"SLIGHTLY" NON-SEPARABLE DATA



Pop Quiz #1

Quick Question!

Why might we need a "soft margin" SVM?

- a) Data is perfectly linearly separable

Pop Quiz #1

Quick Question!

Why might we need a "soft margin" SVM?

- a) Data is perfectly linearly separable
- b) Data has some noise and outliers

Pop Quiz #1

Quick Question!

Why might we need a "soft margin" SVM?

- a) Data is perfectly linearly separable
- b) Data has some noise and outliers
- c) We want smaller margins

Pop Quiz #1

Quick Question!

Why might we need a "soft margin" SVM?

- a) Data is perfectly linearly separable
- b) Data has some noise and outliers
- c) We want smaller margins
- d) To avoid using kernels

Pop Quiz #1

Quick Question!

Why might we need a "soft margin" SVM?

- a) Data is perfectly linearly separable
- b) Data has some noise and outliers
- c) We want smaller margins
- d) To avoid using kernels

Pop Quiz #1

Quick Question!

Why might we need a "soft margin" SVM?

- a) Data is perfectly linearly separable
- b) Data has some noise and outliers
- c) We want smaller margins
- d) To avoid using kernels

Answer: b) Data has some noise and outliers - soft margin allows controlled violations.

Soft-Margin SVM

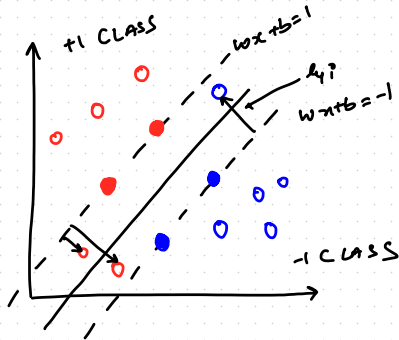
- Can we learn SVM for “slightly” non-separable data without projecting to a higher space?

Soft-Margin SVM

- Can we learn SVM for “slightly” non-separable data without projecting to a higher space?
- Introduce some “slack” (ξ_i) or loss or penalty for samples - allow some samples to be misclassified

"SLIGHTLY" NON-SEPARABLE DATA

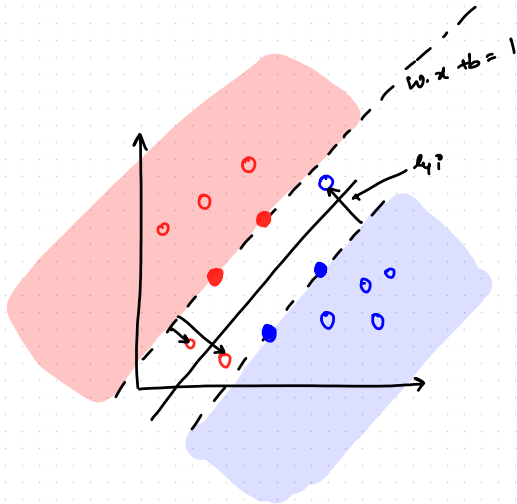
e_{yi} : Distance from margin



ZONE 1

$$y_i = +1$$

$$w \cdot x_i + b \geq 1$$



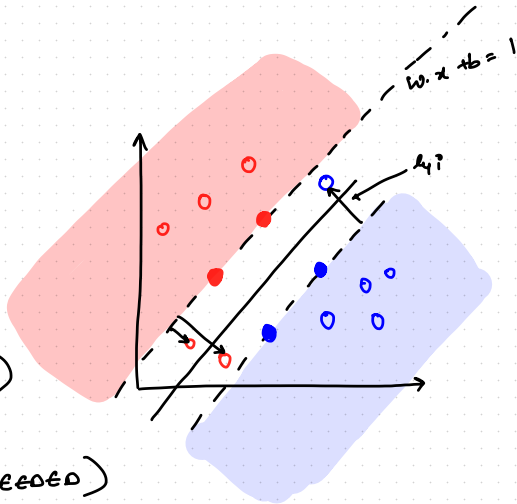
ZONE 1

$$y_i (\vec{w} \cdot \vec{x}_i + b) > 1$$

$$\text{Loss}_i = 0 \quad (\eta_i = 0)$$



NO PENALTY
(OR SLACK NEEDED)

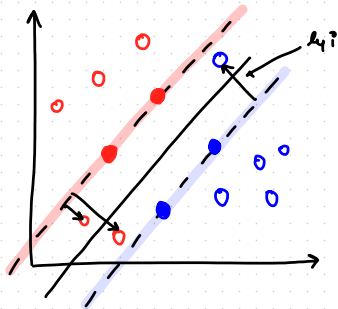


ZONE 2

$$y_i (\vec{w} \cdot \vec{x}_i + b) = 1$$

$$\text{Loss}_i = 0$$

$$(y_i = 0)$$



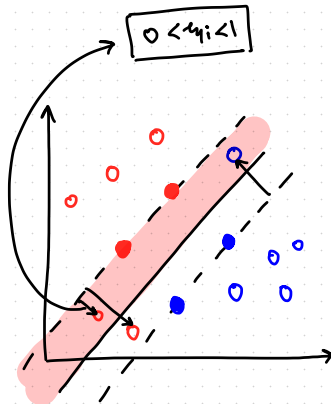
ZONE 3

$$y_i (\bar{w} \cdot \bar{x}_i + b) < 1$$

$$\text{LOSS}_i \neq 0 \quad (0 < \eta_i < 1)$$

POINT CORRECTLY
CLASSIFIED

(BUT WRONG
SIDE OF MARGIN)



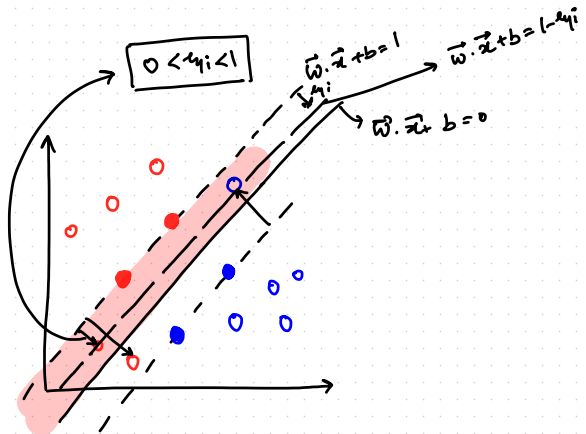
ZONE 3

$$y_i (\vec{w} \cdot \vec{x}_i + b) < 1$$

$$\text{LOSS}_i \neq 0 \quad (0 < \eta_i < 1)$$

POINT CORRECTLY
CLASSIFIED

(BUT WRONG
SIDE OF MARGIN)



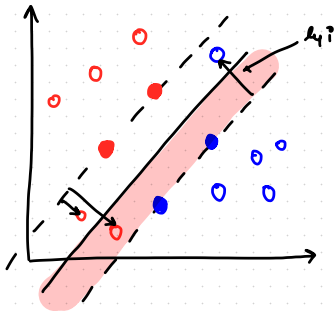
ZONE 4

$$y_i (\vec{w} \cdot \vec{x}_i + b) < 1$$

POINT INCORRECTLY
CLASSIFIED

$$\text{Loss}_i \neq 0$$

$$h_i > 1$$



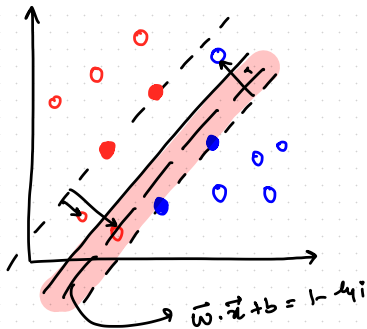
ZONE 4

$$y_i (\vec{w} \cdot \vec{x}_i + b) < 1$$

POINT INCORRECTLY
CLASSIFIED

$$\text{Loss}_i \neq 0$$

$$\eta_i > 1$$



Soft-Margin SVM

Change Objective

$$\begin{aligned} &\text{minimize } \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i \\ &\text{s.t. } y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i \end{aligned}$$

Soft-Margin SVM

Change Objective

$$\begin{aligned} &\text{minimize } \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i \\ &\text{s.t. } y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i \end{aligned}$$

In Dual:

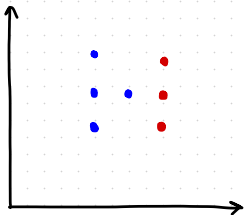
$$\text{minimize } \sum_{i=1}^n \alpha_i - \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j$$

s.t.

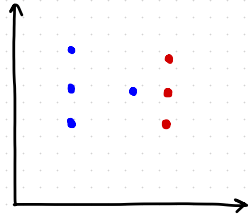
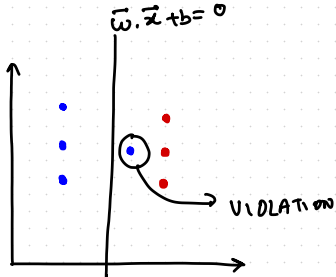
$$0 \leq \alpha_i \leq C \quad \& \quad \sum_{i=1}^n \alpha_i y_i = 0$$

BIAS - VARIANCE

TRADE-OFF

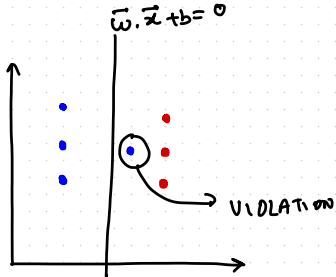


BIAS- VARIANCE TRADE-OFF

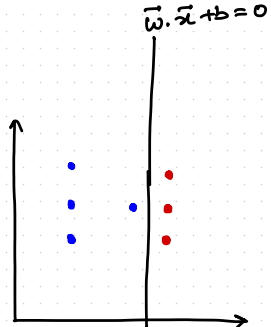


LOW C
 LOW PENALTY FOR VIOLATION
 HIGH TRAIN ERROR
 HIGH BIAS

BIAS- VARIANCE TRADE-OFF



LOW C
 LOW PENALTY FOR VIOLATION
 HIGH TRAIN ERROR
 HIGH BIAS
 BIG MARGIN



HIGH C
 HIGH PENALTY FOR VIOLATION
 HIGH VARIANCE
 SMALL MARGIN

Pop Quiz #2

Quick Question!

What happens when the regularization parameter C is very large?

- a) The model becomes more tolerant to misclassifications

Pop Quiz #2

Quick Question!

What happens when the regularization parameter C is very large?

- a) The model becomes more tolerant to misclassifications
- b) The model tries to classify all training points correctly

Pop Quiz #2

Quick Question!

What happens when the regularization parameter C is very large?

- a) The model becomes more tolerant to misclassifications
- b) The model tries to classify all training points correctly
- c) The margin becomes larger

Pop Quiz #2

Quick Question!

What happens when the regularization parameter C is very large?

- a) The model becomes more tolerant to misclassifications
- b) The model tries to classify all training points correctly
- c) The margin becomes larger
- d) Regularization increases

Pop Quiz #2

Quick Question!

What happens when the regularization parameter C is very large?

- a) The model becomes more tolerant to misclassifications
- b) The model tries to classify all training points correctly
- c) The margin becomes larger
- d) Regularization increases

Pop Quiz #2

Quick Question!

What happens when the regularization parameter C is very large?

- a) The model becomes more tolerant to misclassifications
- b) The model tries to classify all training points correctly
- c) The margin becomes larger
- d) Regularization increases

Answer: b) The model tries to classify all training points correctly - high variance!

Bias Variance Trade-off for Soft-Margin SVM

Low $C \implies$ Higher train error (higher bias)

High $C \implies$ Very sensitive to datasete (high variance)

Soft-Margin SVM

If $C \rightarrow 0$

Objective \rightarrow minimize $\frac{1}{2} \|\mathbf{w}\|^2$

\implies Choose large margin (without worrying for ξ_i s)

$$\text{Recall: Margin} = \frac{2}{\|\mathbf{w}\|}$$

If $C \rightarrow \infty$ (or very large) Objective \rightarrow minimize $C \sum \xi_i$ or
choose \mathbf{w} , b , s.t. ξ_i is small!

Pop Quiz #3

Quick Question!

What is the equivalent of hard margin?

- a) $C \rightarrow 0$

Pop Quiz #3

Quick Question!

What is the equivalent of hard margin?

- a) $C \rightarrow 0$
- b) $C \rightarrow \infty$

Pop Quiz #3

Quick Question!

What is the equivalent of hard margin?

- a) $C \rightarrow 0$
- b) $C \rightarrow \infty$

Pop Quiz #3

Quick Question!

What is the equivalent of hard margin?

- a) $C \rightarrow 0$
- b) $C \rightarrow \infty$

Answer: b) $C \rightarrow \infty$ - No violations allowed!

Pop Quiz #4

Quick Question!

For a support vector with slack variable $\xi_i = 1.5$, this point is:

- a) On the margin boundary

Pop Quiz #4

Quick Question!

For a support vector with slack variable $\xi_i = 1.5$, this point is:

- a) On the margin boundary
- b) Correctly classified but within margin

Pop Quiz #4

Quick Question!

For a support vector with slack variable $\xi_i = 1.5$, this point is:

- a) On the margin boundary
- b) Correctly classified but within margin
- c) Misclassified

Pop Quiz #4

Quick Question!

For a support vector with slack variable $\xi_i = 1.5$, this point is:

- a) On the margin boundary
- b) Correctly classified but within margin
- c) Misclassified
- d) Outside both margins

Pop Quiz #4

Quick Question!

For a support vector with slack variable $\xi_i = 1.5$, this point is:

- a) On the margin boundary
- b) Correctly classified but within margin
- c) Misclassified
- d) Outside both margins

Pop Quiz #4

Quick Question!

For a support vector with slack variable $\xi_i = 1.5$, this point is:

- a) On the margin boundary
- b) Correctly classified but within margin
- c) Misclassified
- d) Outside both margins

Answer: c) Misclassified - since $\xi_i > 1$!

Soft-Margin SVM

Types of support vectors:

- Zone 2: $y_i(\mathbf{w} \cdot \mathbf{x}_i + b) = 1$

\therefore As C increases, # support vectors decreases

Notebook: SVM-soft-margin

Soft-Margin SVM

Types of support vectors:

- Zone 2: $y_i(\mathbf{w} \cdot \mathbf{x}_i + b) = 1$
- Zone 3: $0 < \xi_i < 1$ (correctly classified)

\therefore As C increases, # support vectors decreases

Notebook: SVM-soft-margin

Soft-Margin SVM

Types of support vectors:

- Zone 2: $y_i(\mathbf{w} \cdot \mathbf{x}_i + b) = 1$
- Zone 3: $0 < \xi_i < 1$ (correctly classified)
- Zone 4: $\xi_i > 1$ (Misclassified)

\therefore As C increases, # support vectors decreases

Notebook: SVM-soft-margin

SVM Formulation in the Loss + Penalty Form

Objective:

$$\text{minimize } \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i$$

Now:

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i$$

$$\xi_i \geq 1 - y_i(\mathbf{w} \cdot \mathbf{x}_i + b)$$

But $\xi_i \geq 0$

$$\therefore \xi_i = \max [0, 1 - y_i(\mathbf{w} \cdot \mathbf{x}_i + b)]$$

Pop Quiz #5

Quick Question!

The hinge loss function $\max[0, 1 - y_i(\mathbf{w} \cdot \mathbf{x}_i + b)]$ is:

- a) Convex and differentiable everywhere

Pop Quiz #5

Quick Question!

The hinge loss function $\max[0, 1 - y_i(\mathbf{w} \cdot \mathbf{x}_i + b)]$ is:

- a) Convex and differentiable everywhere
- b) Convex but not differentiable at one point

Pop Quiz #5

Quick Question!

The hinge loss function $\max[0, 1 - y_i(\mathbf{w} \cdot \mathbf{x}_i + b)]$ is:

- a) Convex and differentiable everywhere
- b) Convex but not differentiable at one point
- c) Non-convex but differentiable

Pop Quiz #5

Quick Question!

The hinge loss function $\max[0, 1 - y_i(\mathbf{w} \cdot \mathbf{x}_i + b)]$ is:

- a) Convex and differentiable everywhere
- b) Convex but not differentiable at one point
- c) Non-convex but differentiable
- d) Neither convex nor differentiable

Pop Quiz #5

Quick Question!

The hinge loss function $\max[0, 1 - y_i(\mathbf{w} \cdot \mathbf{x}_i + b)]$ is:

- a) Convex and differentiable everywhere
- b) Convex but not differentiable at one point
- c) Non-convex but differentiable
- d) Neither convex nor differentiable

Pop Quiz #5

Quick Question!

The hinge loss function $\max[0, 1 - y_i(\mathbf{w} \cdot \mathbf{x}_i + b)]$ is:

- a) Convex and differentiable everywhere
- b) Convex but not differentiable at one point
- c) Non-convex but differentiable
- d) Neither convex nor differentiable

Answer: b) Convex but not differentiable at $y_i(\mathbf{w} \cdot \mathbf{x}_i + b) = 1$!

SVM Formulation in the Loss + Penalty Form

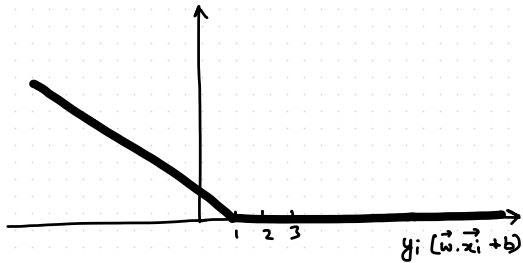
∴ Objective is:

$$\text{minimize } C \sum \xi_i + \frac{1}{2} \|\mathbf{w}\|^2$$

$$\Rightarrow \text{minimize } C \sum_{i=1}^N \max [0, 1 - y_i(\mathbf{w} \cdot \mathbf{x}_i + b)] + \frac{1}{2} \|\mathbf{w}\|^2$$

$$\Rightarrow \text{minimize } \underbrace{\sum_{i=1}^N \max [0, 1 - y_i(\mathbf{w} \cdot \mathbf{x}_i + b)]}_{\text{Loss}} + \underbrace{\frac{1}{2C} \|\mathbf{w}\|^2}_{\text{Regularisation}}$$

HINGE LOSS



Loss Function for Sum (Hinge Loss)

Loss function is $\sum_{i=1}^N \max [0, 1 - y_i(\mathbf{w} \cdot \mathbf{x}_i + b)]$

- Case I $y_i(\mathbf{w} \cdot \mathbf{x}_i + b) = 1$
Lies on Margin: $Loss_i = 0$

Loss Function for Sum (Hinge Loss)

Loss function is $\sum_{i=1}^N \max [0, 1 - y_i(\mathbf{w} \cdot \mathbf{x}_i + b)]$

- Case I $y_i(\mathbf{w} \cdot \mathbf{x}_i + b) = 1$

Lies on Margin: $Loss_i = 0$

- Case II

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) > 1$$

$$Loss_i = 0$$

Loss Function for Sum (Hinge Loss)

Loss function is $\sum_{i=1}^N \max [0, 1 - y_i(\mathbf{w} \cdot \mathbf{x}_i + b)]$

- Case I $y_i(\mathbf{w} \cdot \mathbf{x}_i + b) = 1$

Lies on Margin: $Loss_i = 0$

- Case II

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) > 1$$

$$Loss_i = 0$$

- Case III

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) < 1$$

$$Loss_i \neq 0$$

Hinge Loss Continued

Q) Is hinge loss convex and differentiable?

Convex: ✓

Differentiable: X

Subgradient: ✓

SVM Loss is Convex

Hinge Loss $\sum(\max[0, (1 - y_i(\mathbf{w} \cdot \mathbf{x}_i + b))])$ is convex

Penalty $\frac{1}{2} \|\mathbf{w}\|^2$ is convex

\therefore SVM loss is convex